Technical Report #CS05-10-02 Oregon State University October 5,2005

Getting to the information you already have

Vidya Rajaram, Simone Stumpf, Margaret Burnett, Anton Dragunov, [+]Evelyn Wick, Juliana Lynn, Kevin Johnsrude, Jonathan Herlocker
Oregon State University
Corvallis, Oregon 97331
{rajaramv,stumpf,burnett,anton,lynnju,johnsrud,herlock}@eecs.oregonstate.edu
[+]{perezev@lifetime.oregonstate.edu}

ABSTRACT
Knowledge workers need to find information but even when it is stored on their local computer systems, finding it can be costly. There are many researchers working on solutions to reduce these costs, but there has been little research into exactly what these costs are, and what the ties are between these costs and users' choices between ways to access their local information. This paper provides a methodology for investigating such issues, and reports empirical results on ways of accessing local, task-relevant resources (e.g. document files), their associated costs, and users' sensitivities to certain kinds of costs. Our results fill in gaps in what has been known about the problem, thereby helping to inform research on solutions to the problem.

Author Keywords
accessing resources, user costs, finding information, opening files

ACM Classification Keywords
H.1.2 Models and Principles: User/Machine Systems

H.3.2 Information Storage and Retrieval

H.5.2 Information interfaces and presentation (e.g., HCI): User Interfaces

INTRODUCTION

Computer users who spend considerable time processing information (termed *knowledge workers* in this paper) access a significant amount of digital information in the course of doing their jobs. The type of information they access that is of interest in this paper is the information stored in files and folders (termed *resources*) on the user's computer or local area network.

The task of accessing such resources is often not a simple one, in part because knowledge workers work with information that may not be organized according to their current tasks — or not even be very well organized by any criterion (e.g., Figure 1). Although several research efforts have helped to shed light on user processes for organizing and retrieving information [1, 24, 25], and the difficulties they run into along the way [2, 8, 21, 22, 27], these works have not attempted to quantify the cost of these processes and difficulties.

In response to the reported difficulties users have in accessing their resources, there have been several research efforts to help users better organize and retrieve information [3, 12, 14, 18, 19, 22, 29, 32]. This has led to a proliferation of different approaches to support access to resources.

We would like to help inform these efforts to reduce users' costs by providing insights into how high these costs to the user actually are, and to what extent different aspects of cost relate to the choices users make. To enable this sort of investigation, in this paper we present a cost-centric methodology for investigating issues in accessing resources.

We further present empirical results of a study that uses this methodology. We report on which of the many ways to access resources the users actually employed and their associated costs, whether users' choices seemed to be predicted by users' perceptions of cost, and how perceived, optimal, and actual costs related to each other. In considering these questions, our aim is to help provide a better understanding of the *problem* of local resource access, so as to inform research into possible solutions.

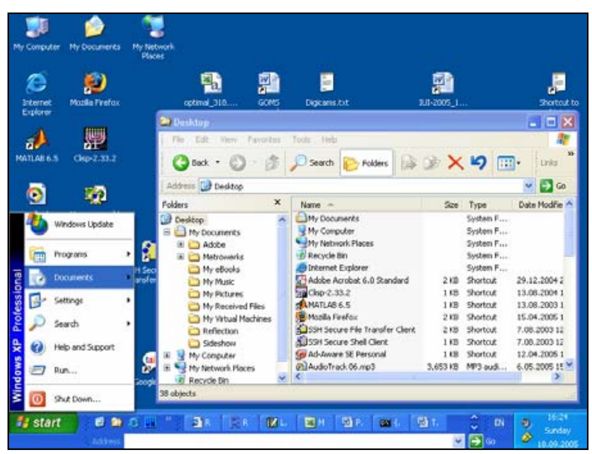


Figure 1. One of our authors' desktop.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **05 OCT 2005**	2. REPORT TYPE	3. DATES COVERED **00-00-2005 to 00-00-2005**

4. TITLE AND SUBTITLE **Getting to the information you already have**	5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Oregon State University,School of Electrical Engineering and Computer Science,1148 Kelley Engineering Center,Corvallis,OR,97331-5501**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Knowledge workers need to find information but even when it is stored on their local computer systems, finding it can be costly. There are many researchers working on solu-tions to reduce these costs, but there has been little research into exactly what these costs are, and what the ties are be-tween these costs and users? choices between ways to ac-cess their local information. This paper provides a method-ology for investigating such issues, and reports empirical results on ways of accessing local, task-relevant resources (e.g. document files), their associated costs, and users? sen-sitivities to certain kinds of costs. Our results fill in gaps in what has been known about the problem, thereby helping to inform research on solutions to the problem.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**24**	

RELATED WORK

Finding Files

Our work builds on the findings of previous researchers who have investigated the preferences and habits of people attempting to access local information, and how well their habits align with the mechanisms available. For example, numerous studies have found that people rely heavily on the classification of information into folders to facilitate retrieval, preferring manual browsing of the file structure over logical search tools [2, 21, 27]. One reason for this may be that users navigate to a desired file in small steps using context as guide [33]. Furthermore, Kaptelinin found that users employed a limited number of manual browsing techniques [23].

These results point to the need for different access methods that may be more appealing to these users. Toward this end, various approaches for personal information management that do not rely on hierarchically structured file systems have been proposed. For example, one organization scheme is the "pile" or "data mountain", a loosely structured information collection [13, 30]. Another approach is to show information in a time-centric visualization [15, 28].

Time-centric organizations have the advantage of allowing the worker to return to records of past work, but for multitasking workers, a single timeline of all activity is a jumble of multiple tasks. The importance of units of work such as projects, activities, and to-do items (which we will call *tasks*) to the knowledge worker has been recognized [1, 17]. Tasks are commonly employed by knowledge workers to organize information they are working on [7]. Various tools have been developed to support grouping information in task-centric workspaces [6, 14, 19, 22, 29, 32].

Further information on the specific costs that users incur can help these efforts to minimize costs. The contribution of this paper, then, is measuring and comparing the users' actual costs and their perceived costs of resource access methods in order to provide a baseline understanding of the problem space. This richer knowledge can inform the context of research into more efficient solutions for local information retrieval.

User Costs

One proposed cognitive model for analyzing decisions is the Attention Investment Model [5], and its cost basis aligns well with our experiment's goals. This predictive model describes how users decide to allocate their attention by conducting a cost-risk-benefit analysis during problem-solving, in which the units they are weighing up are the cost of their attention (closely related to time). The expected benefit of an action is compared to the investment required and to the possible risks that the cost may be paid without receiving the expected benefits after all. The Attention Investment Model will serve as an intellectual framework in our investigation into how people make choices among ways to access resources.

Note that the Attention Investment Model focuses on *perceived* costs, risks, and benefits. This aspect of the model is critical; often, users make choices based on limited information. The perceived amount of benefit includes not only the immediate benefit, but also the extent to which the user believes that their investment of time may reduce future costs. Also, the users' perception of costs, risks, and benefits may be influenced by their past experiences, and may have little to do with the actual costs, risks, and benefits.

Closely related to the Attention Investment Model is the Attentional User Interface (AUI) model [20]. The AUI model also draws from economic models of attention and investment. The AUI model takes into account the information content of a potential interruption by the computer (such as a potential notification of an important incoming email message) in the context of the user's preferences and his/her current and recent online activities as detected by sensors. Although the AUI model does not consider cognitive activities directly, the user's physical activities can be regarded as indirect clues about the user's attention. The AUI model's purpose is to predict whether and how the computer should interrupt the user. From the perspective of our research, this is a critical difference between the AUI model and the Attention Investment Model — the Attention Investment Model considers what the *user* will do under specific conditions, whereas the AUI model considers what the *computer* should do. Thus, the AUI model is not applicable to our goal.

EXPERIMENT DESIGN

To investigate how people access local resources and their reasons for choosing different ways of accessing them, we designed our observational experiment around activities requiring local resources, some of which were required more than once. To reflect the realities of many work environments, we included a mix of interrupted and uninterrupted activities. Extensive electronic transcripts of user interaction events and system status information were recorded automatically by measuring devices and used in our analysis.

Experimental Task

The participants' task was to answer up to 40 questions about the U.S. National Aeronautics and Space Administration (NASA). A question's role was to represent a single activity for a knowledge worker; hence, questions were designed to require thought, information search, and finally information communication.

Each question required information from one or more of the resources (Word and Excel files stored on the participant's computer). In order to maximize the need for these resources, well-known facts about NASA were avoided. The files were named and organized into folders by a non-technical end user who was not involved in the experiment in any other way.

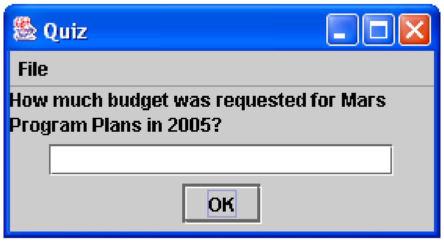


Figure 2. Quiz software: Text input question.

Questions were presented sequentially by software designed for the experiment (Figure 2). This *quiz software* presented the question and an input box for typing in the text answers or a set of radio buttons for answering multiple choice questions. The software prevented participants from skipping questions. The starting question for each participant was chosen at random by the software.

To make the patterns of resource needs as realistic as possible, we designed the questions so that three resources were *key resources*. These files provided answers or partial answers to approximately half the questions in the experiment. This design choice was based on office visits we conducted with 10 knowledge workers (office workers and educators, not software developers) in which we learned that they had a small set of resources, usually two or three, to which they refer repeatedly and frequently during a single activity.

In work environments in which we have particular interest, users have multiple tasks and interruptions. To model this aspect of the real world in our experiment, certain questions (randomly selected prior to the experiment) were interrupted after the participant had opened a predetermined number of resource documents. At that point, the participant was forced to switch to another question (activity). The number of resources triggering a particular question's interruption was also randomly selected prior to the experiment. The participants did not know which questions would be interrupted. Time permitting, the interrupted question was reintroduced later in the experiment.

Once a participant had submitted an answer to a question, Windows Explorer (the graphical file browser) and any open resource documents were closed automatically, so that participants would have to explicitly choose them again if they were needed later. The Word and Excel *applications* themselves, however, were left running, to allow the possibility of using the Recent List in the File menu of the applications or the File Open Dialog box if desired by the participant.

Experimental Procedure

Participants began by filling out a pre-session background questionnaire, and were then led through a brief tutorial. During the tutorial they were instructed to browse through and move resource documents in order to familiarize themselves with the location and contents of the resource documents. By opening the files and looking at their contents, participants also populated some of the most-recently-used lists in the system, particularly My Recent Documents in the Start Menu, and the Recent List in Word and Excel. In the ensuing 15-minute practice session, participants worked with practice questions, which served the dual purpose of getting participants acquainted with the quiz software and of giving them a purposeful method of further familiarizing themselves with the content and location of the resources. Participants then had 45 minutes to perform the main experiment task.

After the 45-minute experiment task, participants filled out a post-session questionnaire, in which they were asked how they usually accessed resources in the Windows environment and what they perceived the cost of each way to be. In order to prevent the questionnaire from influencing participant behavior during the experiment, these questions were included in the post-session questionnaire rather than at the start of the experiment. Although this design choice allows for the possibility of post-rationalization by participants, in this case we considered this to be the less prejudicial choice.

Participants

Participants were recruited by sending announcements to both undergraduate and graduate student mailing lists. The participants were required to have some prior computer experience, including using computers to prepare materials for their studies or work, but computer science majors and students with upper-division coursework in computer science were excluded from participation. Each participant was paid $20. Overall, 39 people (22 women and 17 men) took part in the study, with a mean 3.28 grade-point average (SD=0.49) and a mean 9.09 years of computer experience (SD=3.43).

A COST-CENTRIC METHODOLOGY FOR INVESTIGATING RESOURCE ACCESS

Because we are interested in how users managed their costs in getting to their local resources, we required an investigation methodology based on costs and that also allowed consideration of perceived costs. The methodology we present here is based on the Attention Investment Model for that reason and consists of a model plus three measuring devices. These devices, which are described in detail later, provide data to instantiate the model. They were a measuring device for the ways that participants chose to use, a measuring device for gathering data on other ways that they could have chosen, and a measuring device for gathering data of perceived ways reported by participants.

The Model: Possible Alternatives

The core of our methodology is a state-transition model reflecting all the states and transitions relevant to accessing local resources that we have been able to discover for a standard Windows system. The state-transition model reflects possible paths a user can take through this state space to access a resource, from root (deciding what first step to

use) to leaf (finally opening the desired file). Each transition between nodes thus represents a choice available to the user at that stage as to how to proceed next.

The model explicitly depicts the work required for the user to access a local resource, which can sometimes be quite a lot. For example, consider the path of opening a file using *My Recent Documents* from the *Start* menu (Figure 3). The full diagram depicting the model is too large to include here, but is available at [Appendix A]. In this paper, we use an abbreviation convention, in which each grey node is a *supernode* — a collapsed version of several nodes and transitions. Transitions marked *Decide* denote mental operations, those marked *Select* denote physical user actions such as mouse clicks and keypresses, and those marked *Navigate* are a shorthand notation indicating that both a mental operation to decide which way to proceed and the sequence of physical user actions to reach the next state are included.

In this example, first the user decides how to proceed from the top level, transitioning to the *Start Menu Chosen* state. The user then takes the actions to activate the Start Menu (Select transition). The user then navigates to the My Recent Documents supernode. There are then three more transitions within this supernode required to finally get the file open. At any time, the user could, of course, decide to try a different alternative. This is depicted in the figure by the Decide transition leading out of the largest supernode up to the top level.

We needed to group alternatives that had only trivial differences into classes to make further analysis tractable and useful. These classes will be referred to as *alternatives* in this paper. To perform the grouping, we selected user interface components which play a high-level function in allowing access to resources (in contrast to low-level functional components such as checkboxes) as the granularity that could form the basis of an alternative. That is, each alternative is an initial user interface state paired with a user interface component that is reachable from this initial state and which may lead to a resource. The total set of alternatives we have found that are possible within the Windows environment are termed *possible alternatives*. Not all possible alternatives are available to the user in a given situation (e.g., a resource may not be in the My Recent Documents list); this subset of possible alternatives is termed *available alternatives*. The choices actually made by the user are their *chosen alternatives*.

This model is quite versatile for cost-based investigation into resource access, because it affords a variety of instantiations. For example, it can be used with a GOMS-based [9] approach to sum up the transitions (operators, in GOMS terminology) and the times that have been empirically established for these operators to provide estimates of likely completion times for each alternative.

The model only answers questions about possible alternatives. Our methodology supplements this model-based view by drawing on observational data to instantiate the model

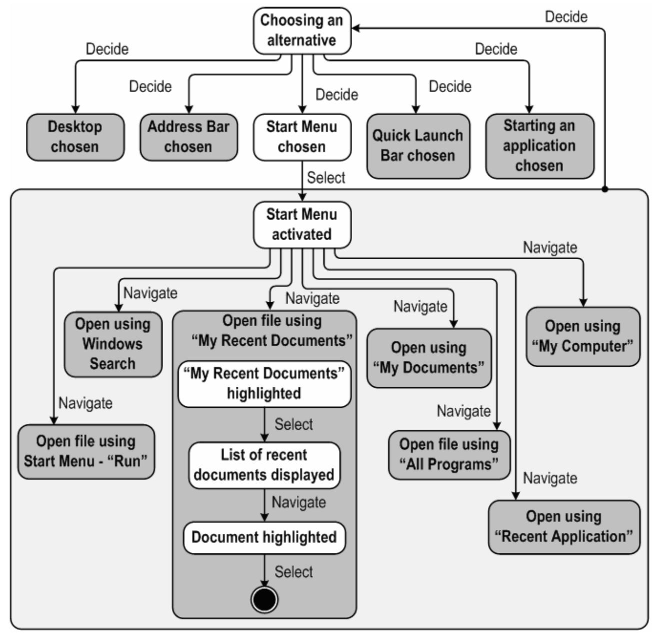


Figure 3. States and transitions involved in opening a file using the Start Menu's My Recent Documents.

with respect to available and chosen alternatives and their costs. Hence, a set of measuring devices is required alongside the model but these may depend on specific settings, such as computing environments, the characteristics of user interaction reporting provided within the environments and the listening capabilities of the measuring device instantiations. We have developed these devices for our experimental setting, as we describe next.

Measuring Device 1: Chosen Alternatives

This measuring device provided data on chosen alternatives and a calculation of their *actual costs*. We decided to use TaskTracer [14] to record each participant's actions and the time needed to complete those actions, generating an electronic transcript of each participant's session. TaskTracer listened in on all user interaction events such as keypresses, mouse clicks, and window focus changes. File open and close events were also recorded. Although the implementation of this capability may seem straightforward, there were a number of technical difficulties, because it involved collecting events from several different applications (Windows, Word, and Excel). These applications have different conventions as to which events they report and how, and substantial implementation effort was required to gather all the relevant events. For this reason, selected transcripts of both pilot participants and the main experiment's participants were manually analyzed to ensure that all relevant alternatives were either reported or deducible from reported data.

In our instantiation of the measuring device, actual costs for individual file accesses were measured in seconds elapsed from the time of the first interaction event in the electronic transcript (state in the model) corresponding to the moment

the participant started looking for a file (i.e., the window focus changed from the quiz software) to the time of the file open event.

Measuring Device 2: Other Available Alternatives

The second measuring device provided data on the “roads not taken” — alternatives to access resources that a participant did *not* use in each situation but that might have been cheaper. The model makes the sequence and number of steps associated with all possible alternatives explicit, but since it does not take the participant’s specific context into account, it cannot alone enumerate which of those possible alternatives are actually available. To gather data on available alternatives, the measuring device recorded *information snapshots* at each time a new question had been posed by the quiz software and each time the participant opened a resource. This information snapshot included extensive system state information, such as files present in My Recent Documents, files present in the *Recent Lists* of each tracked application, files present in *My Documents*, current folder in the *File Open Dialog* box, and so on. The information snapshots were stored alongside the electronic transcript of user behavior.

This measuring device provided data on available alternatives and formed the basis for a calculation of *optimal costs* — the costs of the least expensive alternatives a participant could have chosen to access a particular resource at a particular moment. Given an alternative available in a situation (discerned from the state information captured by the information snapshot), optimal cost was computed as follows: the available alternative’s state in the model was located, and the number of mental operations, keypresses, and mouse clicks to get to it from the initial node were counted. These operations were multiplied by the established times for these operations [9], and then the total times were summed. The achievability of these times for alternatives was then cross-checked, both by comparison to participant times when they actually used these available alternatives, and by asking experienced computer users otherwise not involved in the experiment to open files using these alternatives. In all the cross-checks, the humans in these situations differed from the GOMS-based approach by no more than 20%. When there were differences, the humans were faster, indicating that our optimal cost estimates were conservative. Finally, the alternative with the lowest cost yielded the optimal cost.

Measuring Device 3: Participant-Reported Alternatives

This measuring device enabled the calculation of *perceived costs* and *frequency of previous use*. We employed a post-session questionnaire to gather information about participants’ perceptions of the alternatives (available in [Appendix C]. In particular, the participants were asked 5-value Likert-scale questions about how frequently they used each alternative (never to always) and what their perceived cost of each alternative was (very easy to very hard).

RESULTS

How Many Ways Are There? (Analytical Results)

Accessing a file involves a number of actions, both physical and mental. The presence of mental cost is not surprising, but the number of alternatives to be weighed as part of this mental cost is surprising. Our state-transition model revealed 27 of these alternatives! We do not enumerate all of them here (see [Appendix B] for the complete list), but the ones that our participants actually used will be detailed shortly. Clearly, the cost of making the decision from among so many different alternatives would be high, even if all of these alternatives were known to the user. This suggests that there could be some alternatives that the user never chooses, which we will investigate next.

Which Alternatives Were Used?

The following table identifies all of the alternatives that participants actually used to access files during the experiment (Table 1). Note that only 11 of the 27 possible alternatives were chosen across all participants during the study.

As the table shows, participants overwhelmingly chose to access resources through alternatives involving Windows Explorer. (These results also mirror the proportion of individual file access for which these alternatives were used.)

Alternatives	Code	Number	Percent
Windows Explorer (existing window)	WE	35	89.74
Windows Explorer (Address Bar)	AB	16	41.02
Windows Explorer (My Computer on Desktop)	MC	15	38.46
File Open Dialog (application)	FO	8	20.51
Recent List (application)	RL	7	17.94
Windows Search (Windows Explorer)	WX	7	17.94
Windows Explorer (Start Menu)	WM	6	15.38
Windows Explorer (folder shortcut on Desktop)	WD	3	7.69
Windows Explorer (Start Menu - RUN)	WR	2	5.12
Windows Search (Start Menu)	WS	2	5.12
My Recent Documents (Start Menu)	RD	1	2.56

Number: Number of participants (Total = 39)
Percent: Percentage of participants who used a particular alternative

Table 1. All alternatives used during experiment

More importantly, once they had used Windows Explorer and there was an existing window for it, they kept returning to it (even though it may have been hidden).

Discussion

It seems that the Windows Explorer component was very attractive to participants, which is consistent with previous findings which revealed that participants preferred strategies that gave interactive feedback about their current status towards locating a file [16, 33]. Although all of the alternatives are interactive, Windows Explorer excels in this regard by continually keeping users apprised of the context, which may help to explain its apparent popularity.

Note that some alternatives were available only at certain points. For example, locating a file through the My Recent Documents alternative would only be successful if this resource had been opened recently enough to feature in the list. In the words of the Attention Investment Model [5], attempting such alternatives involved a risk to the participant that they may not receive the benefit of locating the file using that alternative. Further, yet higher cost would be incurred by having to backtrack and choose another alternative.

The alternatives list in Table 1 is useful in another way too, because it validates that the model included in our methodology was well-suited to our purposes: no participant used any alternatives beyond the ones revealed by our model of possible alternatives.

What Alternatives Did Individual Participants Use?

Two competing expectations we had prior to our experiment were that people might make extensive use of many alternatives in attempts to optimize their actual "file open" work, or that they might always stay with a single way in order to optimize their decision making work. However, we were wrong on both counts. Instead, most people used 2 to 3 alternatives (mean=2.64) (see histogram on Figure 4).

The network diagram in Figure 5 shows whether and how often a particular alternative was used in conjunction with another alternative. Nodes in the network represent the 11 alternatives that were used by participants in the experiment. An edge between two nodes indicates that both these alternatives were used in the experiment session by the same participant. The thickness of the edge between two alternatives indicates the number of participants who chose this combination of alternatives.

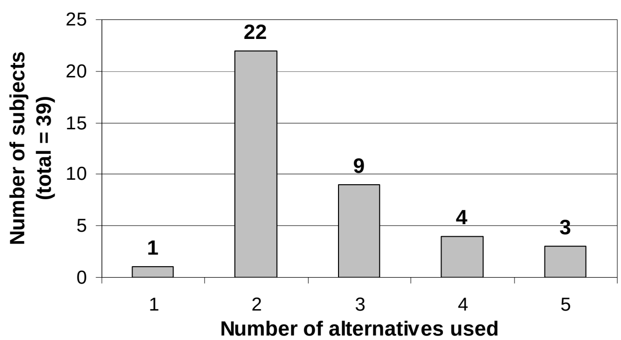
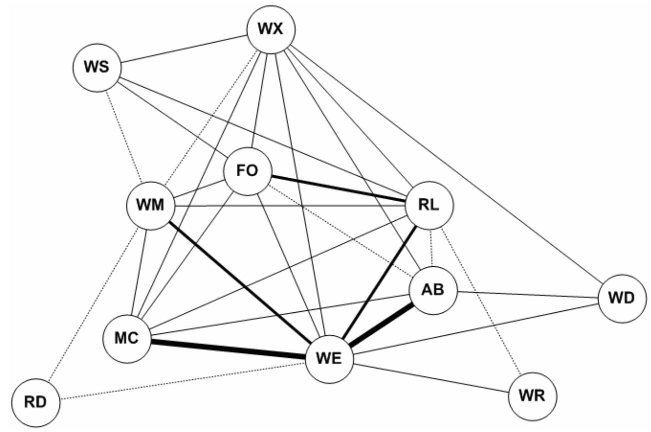


Figure 4. Number of alternatives used by participants.



FO = File Open Dialog (application)
RD = My Recent Documents (Start Menu)
RL = Recent List (application)
AB = Windows Explorer (Address Bar)
WE = Windows Explorer (existing window)
WD = Windows Explorer (folder shortcut on Desktop)
MC = Windows Explorer (My Computer on Desktop)
WM = Windows Explorer (Start Menu)
WR = Windows Explorer (Start Menu - RUN)
WS = Windows Search (Start Menu)
WX = Windows Search (Windows Explorer)

Figure 5. Combinations of alternatives.

Discussion

This network diagram again emphasizes the importance of Windows Explorer to the participants as a core component for accessing resources. Moreover, it reveals an application-centric approach, in that participants tend to use File Open Dialog (FO) together with Recent List (RL) in that application.

In fact, it appears that participants had certain "constellations" of preferred strategies. There were also "anti-constellations" — certain combinations of alternatives that were almost never used together. For example, My Recent Documents (RD) was never used in combination with any kind of Windows Search (WS and WX), nor with any desktop shortcuts (MC and WD), nor with File Open Dialog (FO). The constellations may explain why participants stayed within a small set of alternatives. If so, this in turn suggests that new ways that could be developed may never or rarely be used, if they do not fit well in an existing constellation or catalyze a new one.

What Did Chosen Alternatives Cost?

These descriptions of participant behavior need to be evaluated in light of the reasons for the choices participants made. Participants narrowed down the alternatives that they eventually chose, so what influenced that decision? In Blackwell's terms [5], how did users assess costs in accessing resources?

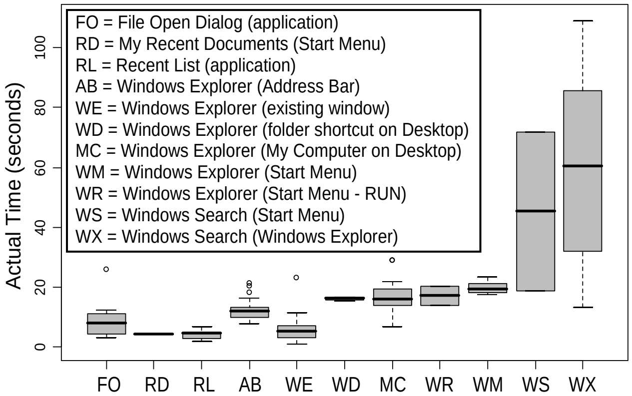


Figure 6. Actual cost for alternatives.

To answer any question about assessment of costs, we first examined the actual cost to the participant of the chosen alternatives (see the Measuring Device 1: Chosen Alternatives section for how actual costs were calculated). The box plot (Figure 6) shows the average actual costs for each alternative. As can be seen from the figure, searching from the Start menu or from Windows Explorer was considerably more costly to the participants than other alternatives.

Discussion

Previous research [21, 23, 33] has indicated that users prefer manual browsing to searching, and the actual costs of these alternatives in our experiment were much lower (and less variable) than alternatives that involved searching. It may be that searching was not used due to lack of experience with computers, as it has been suggested that expertise is a critical factor for using logical search [27]. We therefore asked ourselves whether more experienced participants would be able to reduce their costs by making better choices in their alternatives.

Is Computer Experience a Predictor For Reducing Actual Costs?

As a metric for the prior computer experience we used the number of years a person has been using computers, calculated from the questionnaire data as the number of years of "personal use", or, if this information was not given, the combined years of use in "high school" and "college". We excluded one obvious outlier from our analysis. The outlier was a person who worked with computers for over 30 years, whereas the rest of the subjects had no more than 15 years of experience. The statistical tests showed that there was no significant predictive relationship between prior computer experience and the average actual cost (linear regression, p-value=0.3757, F[1,36]=0.8044, R^2=0.02186).

Hence, we can state that computer experience had no relation to the performance of accessing resources (i.e., participants with more computer experience were not quicker than participants with less computer experience). It therefore appears that greater computer experience did not lead to choosing better alternatives.

Consideration of Optimal Costs

Are participants making choices between alternatives based on optimal costs? To investigate this, we compared the actual costs that participants incurred for each situation in which they needed to access a resource with the optimal costs to access the resource in the same situation (see the Measuring Device 2: Other Available Alternatives section for how optimal costs were calculated).

Our analysis showed that there was no significant relationship between the optimal and the actual costs (linear regression, p-value = 0.4563, F[1,37] = 0.5667, R^2=0.01509). Participants only hit optimal 38% of the time Most of the time, participants seemed to use alternatives even when those alternatives were not the optimal ones. For example, Recent List and My Recent Documents, when they were available, could get participants to the files in only a small number of mouse clicks but instead other, costlier alternatives, such as Windows Explorer, were used.

Discussion

What is becoming clear through our results is that optimal costs and actual costs were not related; participants made choices that were not linked to efficiency.

We pause here to point out the benefit in using our methodology to investigate the predictive nature of costs: it provides a systematic way of evaluating participants' choices by comparing what they chose and what they should have chosen instead, had their decisions been based only on optimal costs. This permitted analysis based on what the optimal alternatives actually were, instead of less precise (researcher intuition-based) notions of which alternatives would be superior.

Especially when coupled with the other results we have presented up to this point, the lack of relationship between optimal and actual costs implies that there were other reasons influencing the choice of alternatives more strongly than an assessment of optimal cost. The Attention Investment Model suggests that *perceived* costs, risks, and benefits play a major role in making choices and that these may be influenced by users' past experiences. Therefore, we decided to investigate the relationships between costs that were participant-reported and the choices that they made.

The Role of Participants' Perception

We investigated whether participants' past experience influenced their choice of alternatives by performing logistic regression. This statistical technique, which is based on probabilities expressed as odds, assessed if perceived costs or frequency of previous use were predictive of whether an alternative was selected during the experiment. We found that neither variable (nor interaction between them) was predictive of the choice of an alternative, *with the exception of Address Bar*. For this alternative, frequency of previous use was predictive of selecting it (logistic regression, p-value = 0.0150). The odds that Address Bar would be chosen during the experiment by participants with high re-

ported frequency of usage were 4.25 times the odds that it would be used by those with low usage (95% confidence interval for the odds ratio is 1.23 to 13.65).

Discussion

The finding that perceived costs and frequency of previous use bear no relation to choice of alternative, at first glance, seems to contradict what has been previously assumed. For example, it has been suggested that people use inefficient procedures because they have frequently used these procedures in the past [16]. Furthermore, the Attention Investment Model focuses on the importance of participants' perceptions. However, we suggest that what our result reveals is participants' particular sensitivity to learning new alternatives for accessing resources.

Evidence for this comes from the fact that the only time that previous frequency of use had any bearing on whether that alternative would be selected was in the case of Address Bar, which is a component of Windows that is not visible by default. What we have not yet considered is that there is another cost to participants that may have influenced their decision to choose certain alternatives: the cost of investing in learning new alternatives.

Investment In Learning As a Reason For Decision

As we have just shown, participants were strongly inclined toward some "favorite" set of alternatives for accessing their resources, and stayed within that set of favorites. These results agree with other researchers who have reported a similar phenomenon. For example, Carroll and Rosson have pointed out that the users often use inefficient strategies even when more efficient strategies exist [10], which they explain through the "paradox of the active user." The essence of this paradox is that users often do not spend the time needed to learn new features because they want to get their work done efficiently. Yet, if they spent the time to learn new features, they might get their work done more efficiently. Furthermore, users may attempt to apply prior knowledge to every new situation, thereby avoiding the cost of additional learning when possible. Similar to Carroll and Rosson, Bhavnani and John [4] observed the usage of inefficient strategies even by experts who valued efficiency. Not spending time to learn is a common theme in these prior studies. This is also in keeping with Blackwell's Attention Investment model, which models such decisions as cost-risk-benefit tradeoffs.

The Attention Investment Model suggests that users will be deterred by the cost of learning new features, even when they realize that there may be better ways to access their resources. This kind of cost is termed an *investment* in the model, because ideally it will produce repeated long-term benefits. Further, as with all investments, there is a risk of low or nonexistent return (such as when a feature is not helpful or the user doesn't succeed in learning), thus making the learning investment unattractive. On the post-session questionnaires from the current study, several participants specifically alluded to prior or expected learning costs and risks. For example:

Q: Why do you usually use these ways <the ways you usually use>?

> *"Because that is how I learned and it is easy."*
> *"Because I am not familiar with other ways."*
> *"Easy to use."*

The responses we obtained suggest satisfaction with the current cost-risk-benefit ratio of prior investment in learning, implying no need for additional learning investment. The Attention Investment Model further suggests that users would be unlikely to take the time to mentally compare a large number of alternatives, since the tradeoff between time savings gained from performing an optimal alternative and the time spent in advance on mental calculations to locate that alternative do not seem worthwhile (and our findings replicate this). These various cost and risk factors sum up to a significant obstacle. Overcoming this obstacle is a challenge to any new technique aimed at helping users to access a desired resource.

Behavior Modification Is Possible

Despite the difficulties imposed by this obstacle, our data revealed some surprising evidence that it is possible to overcome it. In designing this experiment, we chose to make Address Bar visible so that participants could use that alternative if they wanted (Address Bar is not visible by default in Windows). The result of this was that Address Bar was visible at the bottom of the screen with its initial contents pointed to a folder on a local disk drive.

In our study, 16 participants (41.02%) used Address Bar (available in the Windows Task Bar), even though that alternative was the least popular way of accessing resources by two different measures from the post-session questionnaire (Figure 7). This is particularly interesting in light of the fact that 5 people out of those 16 reported that they *did not even know* that this alternative existed prior to the experiment. Clearly, for these 5 people, an incentive was present and that incentive was powerful enough to overcome resistance to new learning costs — and modified their behavior.

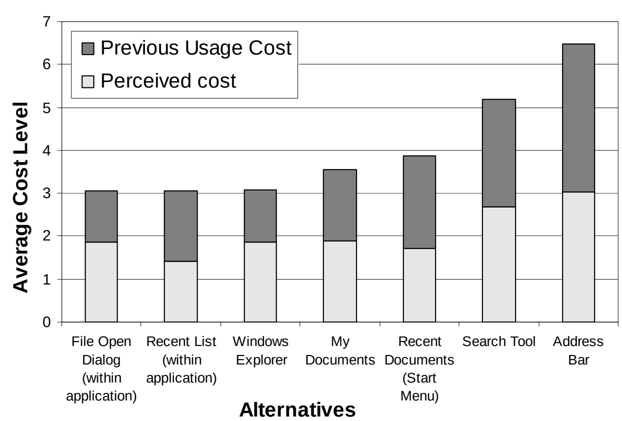


Figure 7. Popularity of Alternatives.

Discussion

Surprise-Explain-Reward [34] is a strategy that has been successfully used in software to attract users' attention to unfamiliar features. It attempts to arouse the user's curiosity about a potentially surprising new feature; he or she can then pursue an interest in the feature by invoking minimalist forms of explanation [11] and, potentially, reap rewards. The way Address Bar was present in the experiment's environment was an instantiation of Surprise-Explain-Reward, although the designers of the bar probably did not realize this. The surprise was Address Bar, which some of these participants had never seen before. The minimalist support available was in the form of an editable example, and because it was in the form of a working example, reading and experimenting with that example were low-cost, low-risk operations. Finally, for some, the experiments paid off: the necessary file was accessed, hence delivering a reward. Once they had this positive experience, they continued to use it throughout the experiment.

This is in keeping with previous successful uses of the Surprise-Explain-Reward strategy [34, 31], suggesting that making a low-cost high-reward alternative visible can motivate users to employ it, when they otherwise would not have even realized it existed.

CONCLUSION

In this paper, we have presented the results of our investigation into how people access resources stored on their local computer systems. One of the contributions of this paper is a methodology for investigating approaches and costs of accessing local resources, including a detailed model of possible alternatives and cost structures.

Furthermore, we described three measuring devices that were motivated by our methodology, which we then instantiated to measure alternatives and costs of resource access: a measuring device for collecting chosen alternatives and their cost, a measuring device for available alternatives and optimal cost, and a measuring device for participant-reported alternatives and costs.

Our study's results showed that participants' choices as to the alternative to access a resource were not related to either the optimal or the perceived cost of that alternative. There are 27 alternatives to consider, and it would be time consuming for a user to consider them all. Instead, our participants chose only a few alternatives (typically 2 or 3). Our findings revealed participants' strong preference for a limited number of alternatives that have low risk (in being able to succeed in reaching the file) and incremental reward (feedback along the way). The cost and risk of learning new alternatives appeared to have outweighed participants' sense of the cost and benefit improvements of actually using them. Hence, another of our contributions is a data point for the Attention Investment Model, in that, in a situation involving a cost-risk-benefit analysis, participants' perceived cost of actually using the alternatives appeared to be outweighed by the perceived cost of learning an alternative. However, it was possible to modify their behavior from this pattern, attracting them to lower-cost alternatives through an instantiation of the Surprise-Explain-Reward strategy. A critical aspect seems to be to alert users to the presence of an alternative that appears to have a low *learning* cost and risk, and which turns out to also deliver low actual cost.

These results reveal that implicit assumptions that may underlie some of the research into improving these costs do not hold. In particular, since our participants considered only a small fraction of the 27 alternatives possible, it seems unlikely that users will voluntarily consider such research's emerging 28th alternative. Thus, three challenges must be faced by research into alternatives that reduce resource access costs for users: (1) how to make users aware of these new alternatives in a way that does not *increase* their already too-high overhead costs; (2) how to communicate low learning cost and low risk of new alternatives — so low that the user will be willing to try them; and (3) how to maximize the chances that a user's initial experimentation with new alternatives will indeed deliver on its promise of low usage cost, so that the user's learning investment will pay off and provide an incentive to continue to use new alternatives.

ACKNOWLEDGMENTS

We thank the participants of our study. This work was supported in part by the EUSES Consortium via NSF grant ITR-0325273, by NSF grant IIS-0133994, and by DARPA grant HR0011-04-1-0005, contract NBCHD030010.

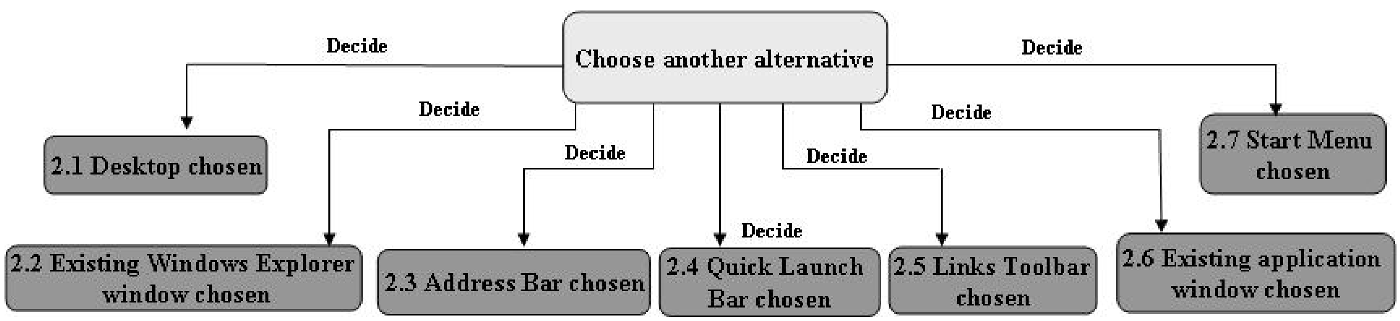
REFERENCES

1. Bannon, L., Cypher, A., Greenspan, S., Monty, M. Evaluation and analysis of users' activity organization. *Proc. CHI 1983*, ACM Press (1983), 54–57.
2. Barreau, D., Nardi, B. Finding and reminding: File organization from the desktop. *SIGCHI Bulletin, 27(3)*, ACM Press (1995), 39–43.
3. Belloti, V., Ducheneaut, N., Howard, M., Smith, I. Taking email to task: The design and evaluation of a task management centered email tool. *Proc. CHI 2003*, ACM Press (2003), 345–352.
4. Bhavnani, S.K., John, B.E. (2000). The strategic use of complex computer systems. *Human–Computer Interaction*, 15(2-3), 107–137.
5. Blackwell, A. First steps in programming: A rationale for attention investment models. *Proc. IEEE Human-Centric Computing Languages and Environments* (2002), 2–10.
6. Boardman, R., Spence, R., Sasse, M.A. Too many hierarchies? The daily struggle for control of the workspace. *Proc. HCI International 2003*, (1), 616–620.
7. Bondarenko, O., Janssen, R. Documents at hand: Learning from paper to improve digital technologies. *Proc. CHI 2005*, ACM Press (2005), 121–130.

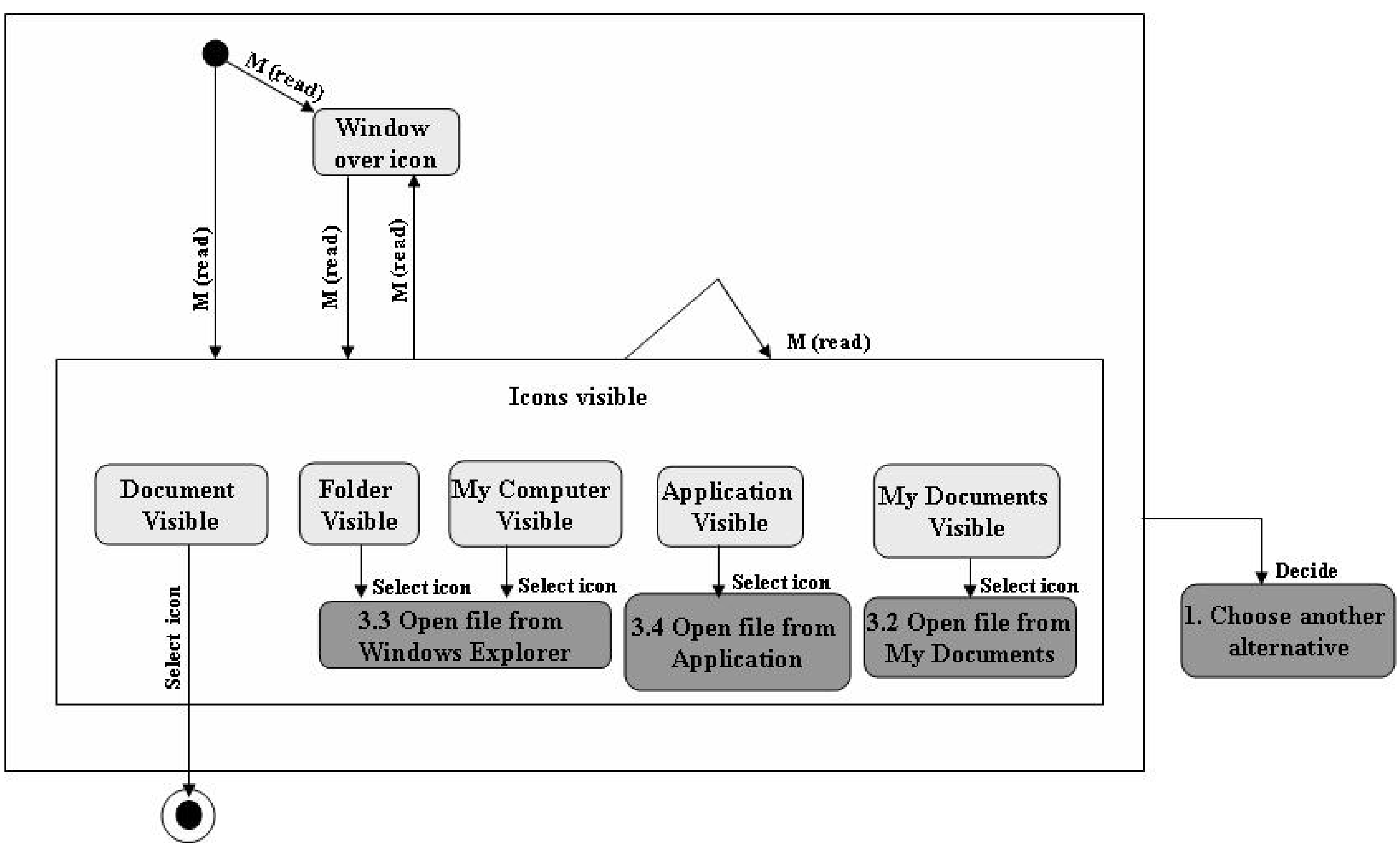
8. Bruce, H., Jones, W., Dumais, S. Information behaviour that keeps found things found. *Information Research*, 10(1), 2004.
9. Card, S.K., Moran, T.P., Newell, A. The keystroke-level model for user performance with interactive systems, *Communications of the ACM*, 23(7), ACM Press (1980), 396–410.
10. Carroll, J.M., Rosson, M.B. Paradox of the active user. In J. M. Carroll (Ed.), *Interfacing thought: Cognitive aspects of human–computer interaction*. MIT Press (1987), 80–111.
11. Carroll, J.M. The Nurnberg Funnel: Designing Minimalist Instruction for Practical Computer Skill, MIT Press (1990).
12. Czerwinski, M., Horvitz, E., Wilhite, S. A diary study of task switching and interruptions. *Proc. CHI 2004*, ACM Press (2004), 175–182.
13. Dourish, P., Edwards, K.W., LaMarca A., Salisbury, M. Presto: An experimental architecture for fluid interactive document spaces. *ACM Transactions on Computer-Human Interaction 1999*, 6(2), ACM Press (1999), 133–161.
14. Dragunov, A.N., Dietterich, T.G., Johnsrude, K., McLaughlin, M., Li, L., Herlocker, J.L. TaskTracer: A desktop environment to support multi-tasking knowledge workers. *Proc. IUI 2005*, ACM Press (2005), 75–82.
15. Freeman, E., Gelernter, D. Lifestreams: A storage model for personal data. In *SIGMOD Record*, 25(1), ACM Press (1996), 80–86.
16. Fu, W., Gray, W. D. (2004). Resolving the paradox of the active user: Stable suboptimal performance in interactive tasks. Cognitive Science, 28(6), 901-937
17. Gonzalez, V., Mark, G. "Constant, constant, multi-tasking craziness": managing multiple working spheres. *Proc. CHI 2004*, ACM Press (2004), 113–120.
18. Gwizdka, J. TaskView: Design and evaluation of a task-based email interface. *Proceedings of the 2002 conference of the Centre for Advanced Studies on Collaborative research*. IBM Press, 2002.
19. Henderson, A., Card, S. Rooms: The use of multiple virtual workspaces to reduce space contention in a window-based graphical user interface. *ACM Transactions on Graphics 1986*,5(3), ACM Press (1986), 211–243.
20. Horvitz, E., Kadie, C., Paek, T., Hovel, D. Models of attention in computing and communication: from principles to applications. *Communications of the ACM*, 46(3), ACM Press (2003), 52–59.
21. Jones, W., Phuwanartnurak. A.J., Gill, R., Bruce, H. Don't take my folders away! Organizing personal information to get things done. In *Ext. Abstracts CHI 2005*, ACM Press (2005), 1505–1508.
22. Kaptelinin, V. UMEA: Translating interaction histories into project contexts. *Proc. CHI 2003*, ACM Press (2003), 353–360.
23. Kaptelinin, V. Creating computer-based work environments: An empirical study of Macintosh users. *Proc. ACM SIGCPR/SIGMIS 1996*, ACM Press (1996), 360–366.
24. Kidd, A. The marks are on the knowledge worker. *Proc. CHI 1994*, ACM Press (1994), 186–191.
25. Malone, T. W. How do people organize their desks? Implications for the design of office information systems. *ACM Transactions on Information Systems 1983*,1(1), ACM Press (1983), 99–112.
26. <details removed for anonymous review>, Technical Report #CS05-10-02, 2005, (available on-line).
27. Ravasio, P., Schar, S., Krueger, H. In pursuit of desktop evolution: User problems and practices with modern desktop systems. *ACM Transactions on Computer-Human Interaction 2004*,11(2), ACM Press (2004), 156–180.
28. Rekimoto, J. Time-machine computing: A time-centric approach for the information environment. *Proc. UIST 1999*, ACM Press (1999), 45–54.
29. Ringel, M. When one isn't enough: an analysis of virtual desktop usage strategies and their implications for design. In *Ext. Abstracts CHI 2003*, ACM Press (2003), 262–263.
30. Robertson, G., Czerwinski, M., Larson, K., Robins, D.C., Thiel, D., von Dantzich, M. Data Mountain: Using spatial memory for document management. *Proc. UIST 1998*, ACM Press (1998), 153–162.
31. Ruthruff, J., Phalgune, A., Beckwith, L., Burnett, M., Cook, C. Rewarding 'good' behavior: End-user debugging and rewards. *Proc. IEEE Visual Languages and Human-Centric Computing* (2004), 115–122.
32. Smith, G., Baudisch, P., Robertson, G., Czerwinski, M., Meyers, B., Robbins, D., Andrews, D. GroupBar: The taskbar evolved. *Proc. OZCHI 2003*.
33. Teevan, J., Alvarado, C., Ackerman, M.S., Karger, D.R., The perfect search engine is not enough: a study of orienteering behavior in directed search. *Proc. CHI 2004*, ACM Press (2004), 415–422.
34. Wilson, A., Burnett, M., Beckwith, L., Granatir, O., Casburn, L., Cook, C., Durham, M., Rothermel, G. Harnessing curiosity to increase correctness in end-user programming. *Proc. CHI 2003*, ACM Press (2003), 305–312.

APPENDIX A – STATE TRANSITION DIAGRAMS

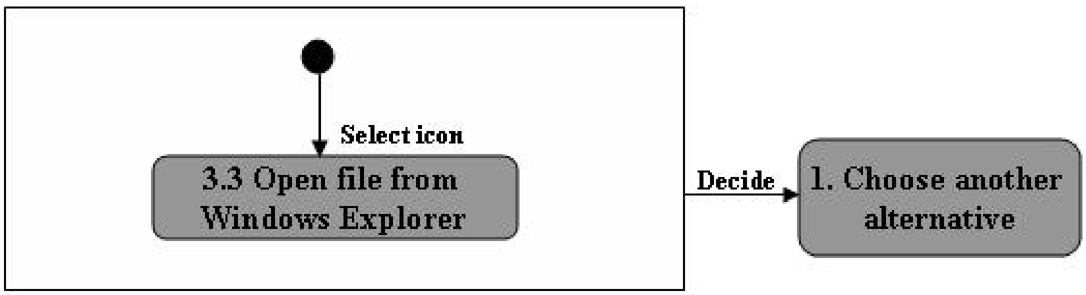
1. Choose another alternative



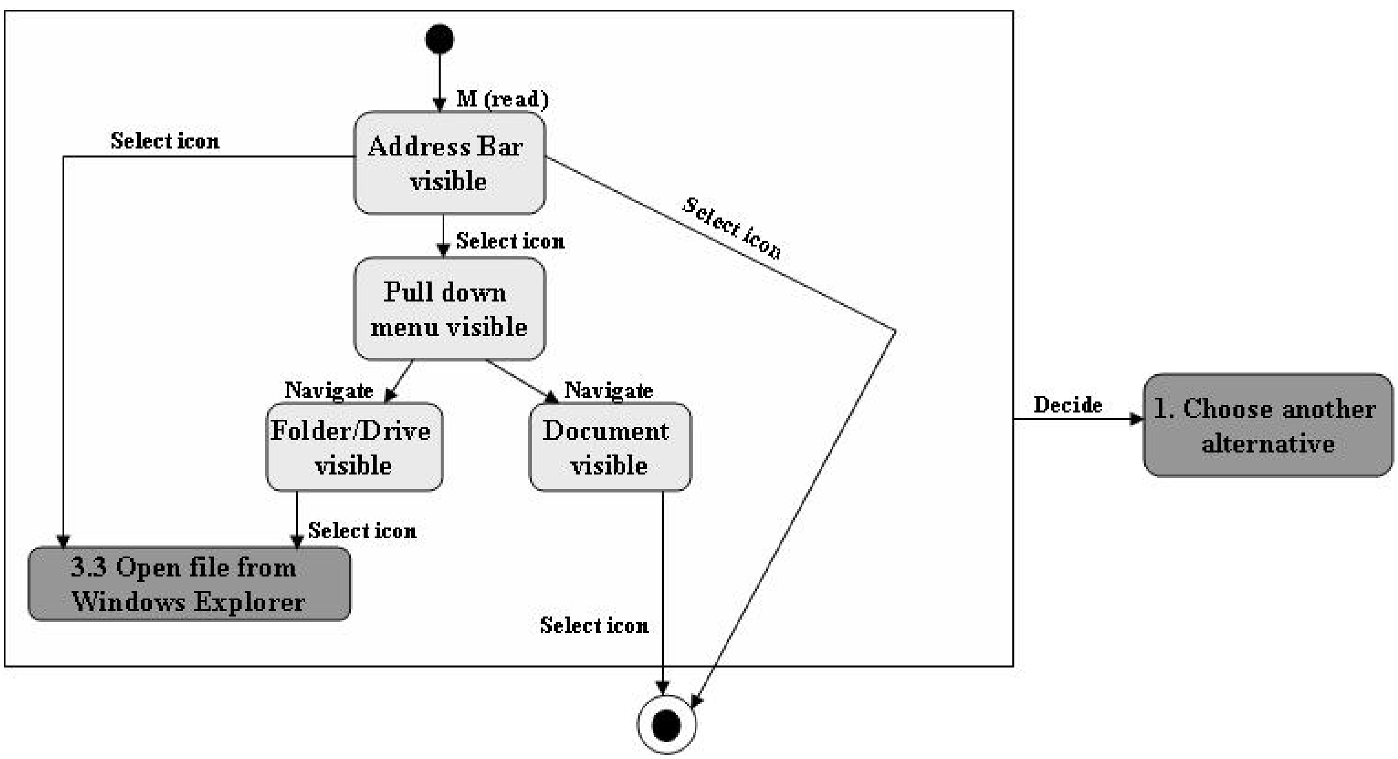
2.1 Desktop chosen



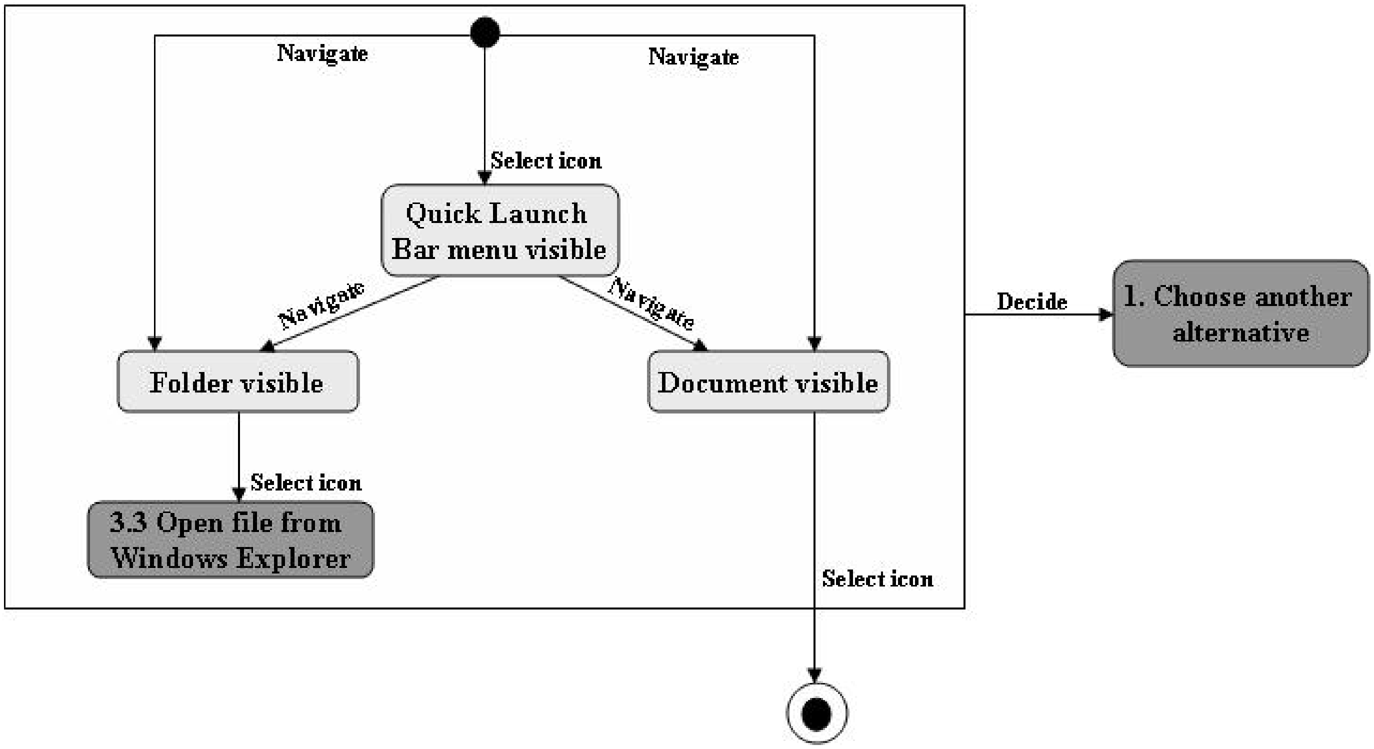
2.2 Existing Windows Explorer chosen



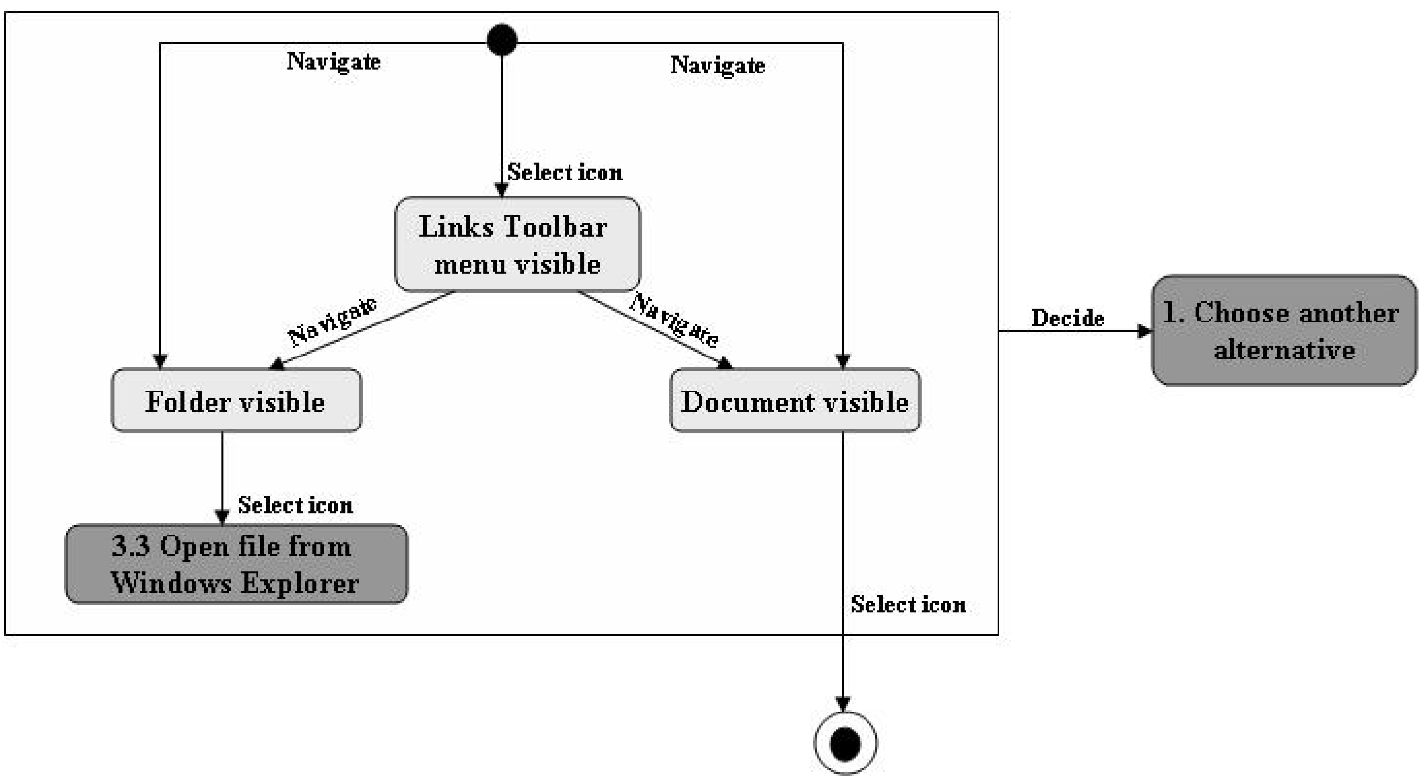
2.3 Address Bar chosen



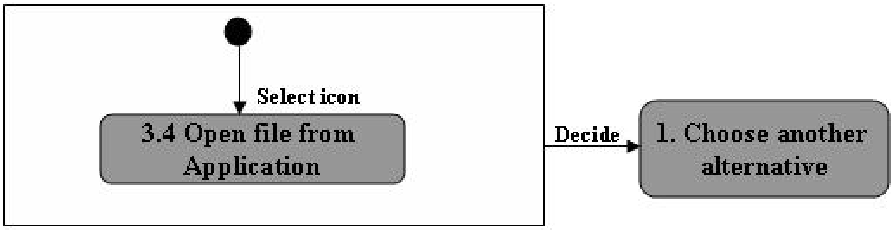
2.4 Quick Launch Bar chosen



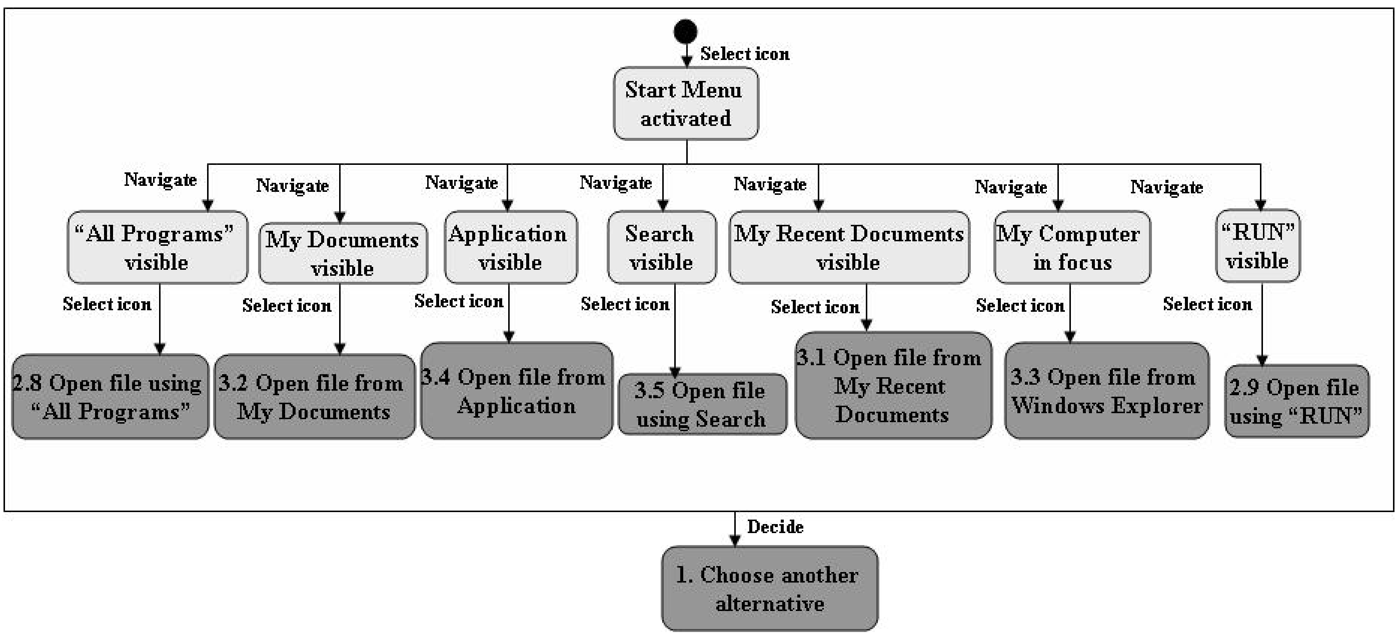
2.5 Links Toolbar chosen



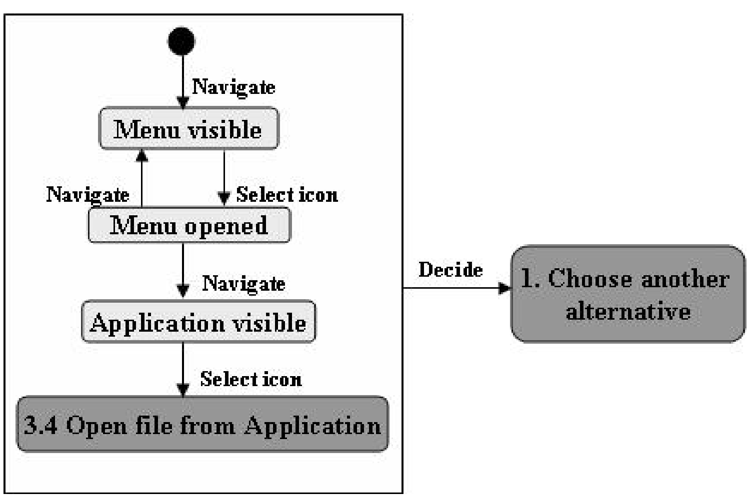
2.6 Existing Application Window chosen



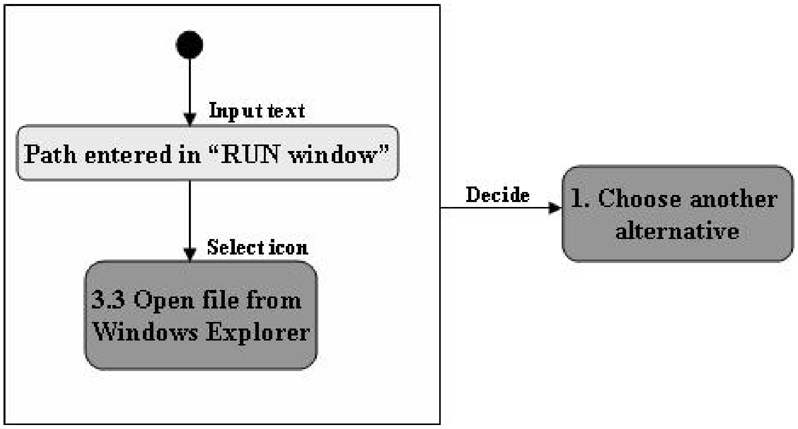
2.7 Start Menu chosen



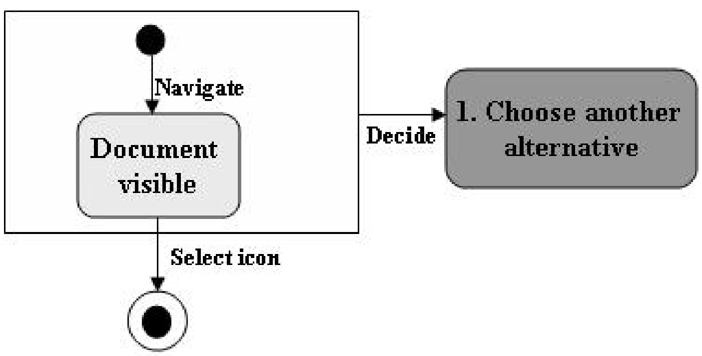
2.8 Open Application from "All Programs"



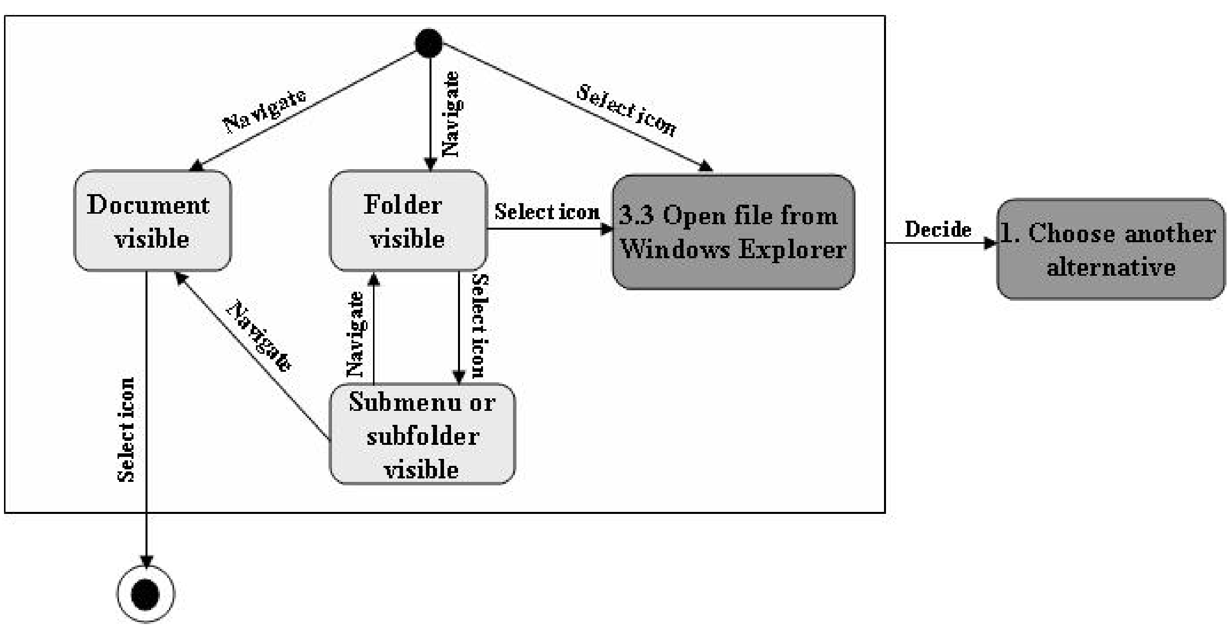
2.9 Open file from "RUN"



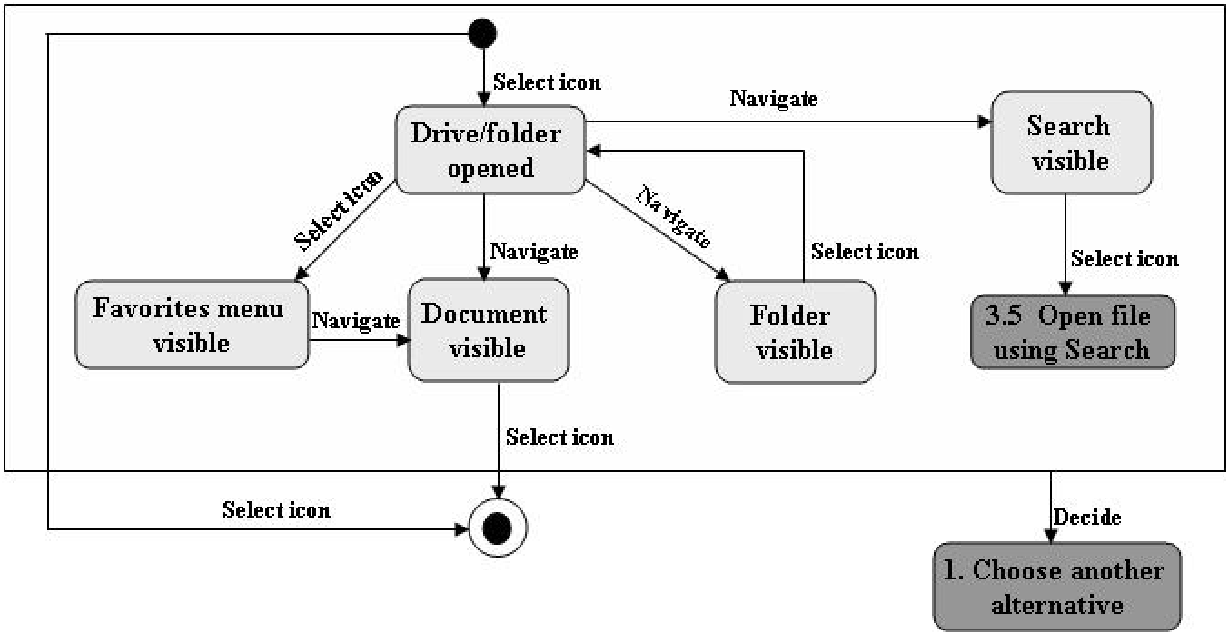
3.1 Open file from My Recent Documents



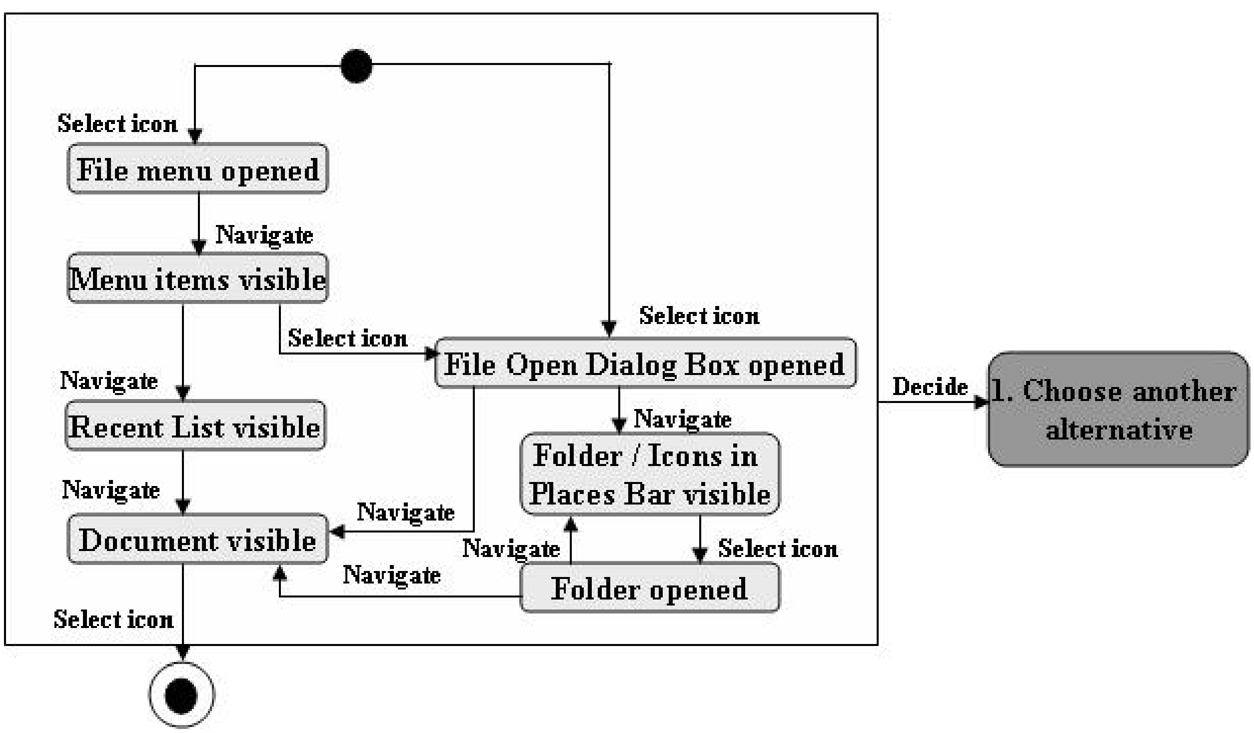
3.2 Open file from My Documents



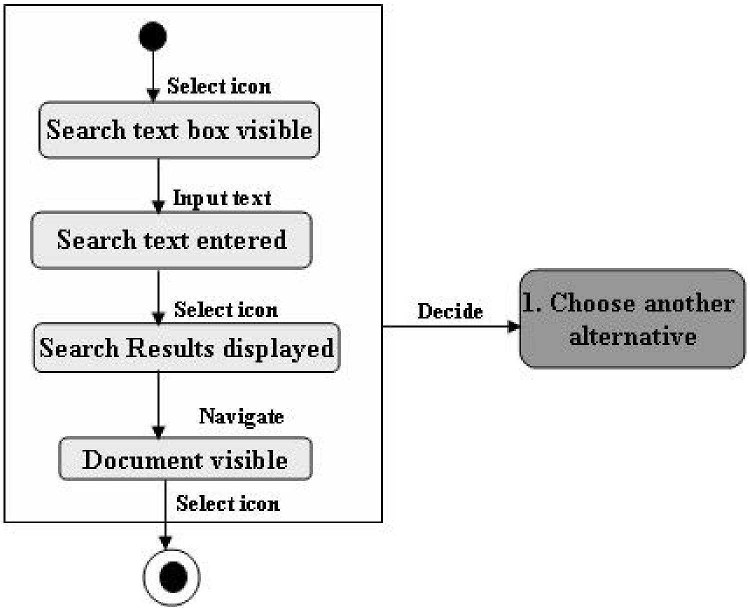
3.3 Open file from Windows Explorer



3.4 Open file from Application



3.5 Open file using Search



Action Grammar

Decide – M

Navigate – (MP)*

Select icon – L | L E | R (P* L | K*) | D | K* (D | E | L | L E | R P* L | R K*)

Input text – K*

Key

M – Mental operation

P – Mouse movement

L – Left-click

E – Enter

R – Right click

K – Key press

D – Double click

APPENDIX B – 27 WAYS TO GET TO A FILE

	Alternative	
Number	**User interface state**	**User interface component reachable from this state**
1	My Documents from Desktop	Click through folders
2	My Documents from Desktop	Choose file directly
3	My Computer from Desktop	Windows Explorer
4	Open new Application from Desktop	Recent List
5	Open new Application from Desktop	File Open Dialog
6	Folder shortcut from Desktop	Windows Explorer
7	Desktop	Choose file directly
8	My Computer from Start Menu	Windows Explorer
9	RUN from Start Menu	Windows Explorer
10	My Documents from Start Menu	Click through folders
11	My Documents from Start Menu	Choose file directly
12	Open application from Start Menu	Recent List
13	Open Application from Start Menu	File Open Dialog
14	Start Menu	Search
15	Start Menu	My Recent Documents
16	Address Bar	Windows Explorer
17	Address Bar	Choose file directly
18	Open folder from Quick Launch	Windows Explorer
19	Quick Launch	Choose file directly
20	Open folder from Links Toolbar	Windows Explorer
21	Links Toolbar	Choose file directly
22	Existing Windows Explorer	Choose file directly
23	Existing Windows Explorer	Use Favorites menu
24	Existing Windows Explorer	Search
25	Existing Windows Explorer	Click through folders
26	Existing Application	Recent List
27	Existing Application	File Open Dialog

APPENDIX C – POST SESSION QUESTIONNAIRE

Subject Number:

POST SESSION QUESTIONNAIRE

1. Please check any one of the following:

LOW HIGH

MENTAL DEMAND

How much mental and perceptual activity (e.g thinking, deciding, calculating, remembering, looking, searching etc.) was required to answer the questions?

TEMPORAL DEMAND

How much time pressure did you feel due to the rate or pace at which the questions appeared?

PERFORMANCE

How successful do you think you were in answering the questions?

EFFORT

How hard did you have to work (mentally or physically) to accomplish your level of performance?

FRUSTRATION LEVEL

How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel while answering the questions?

2. How do you usually get to files in your system?

3. Besides the ones you listed above, do you ever use any other ways? If yes, what are they?

4. Why do you usually use the ways you gave in Question 2?

5. What helps you remember where your files are in the system?

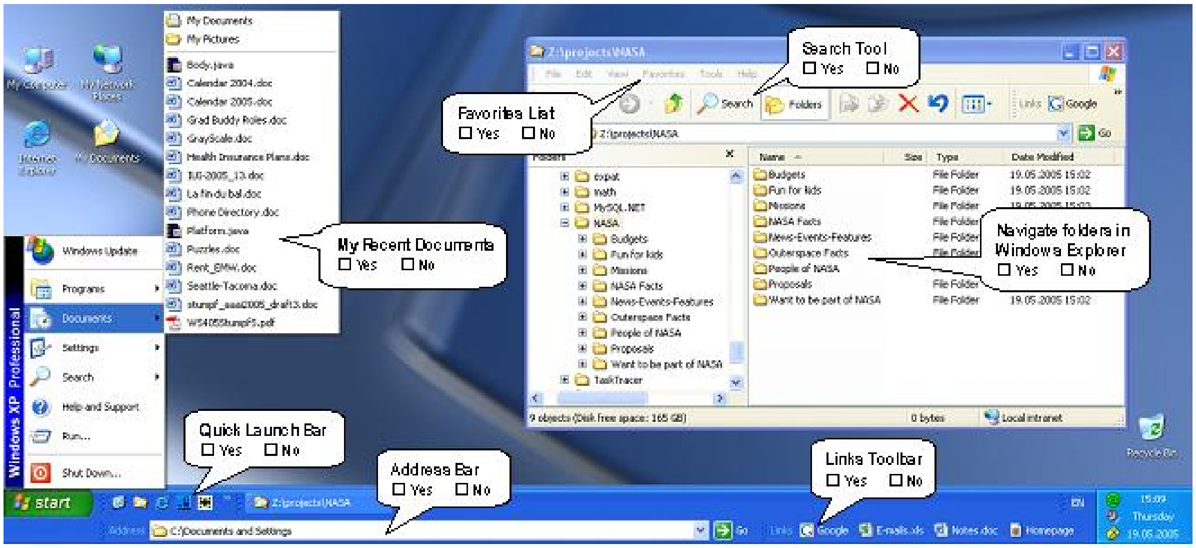
6. Which was more difficult? (Choose one)

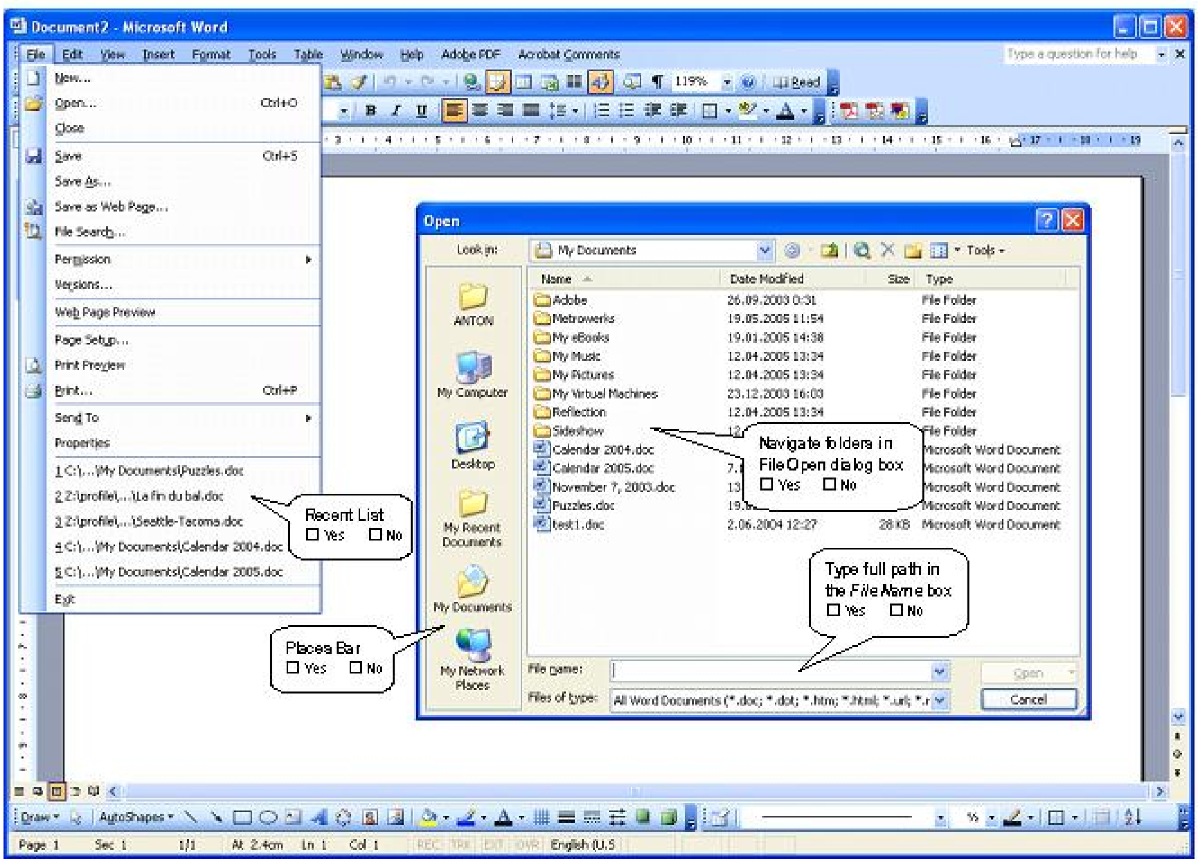
- ☐ Getting the right file
- ☐ Finding the information within the file

7. How do you store files that belong to more than one topic?

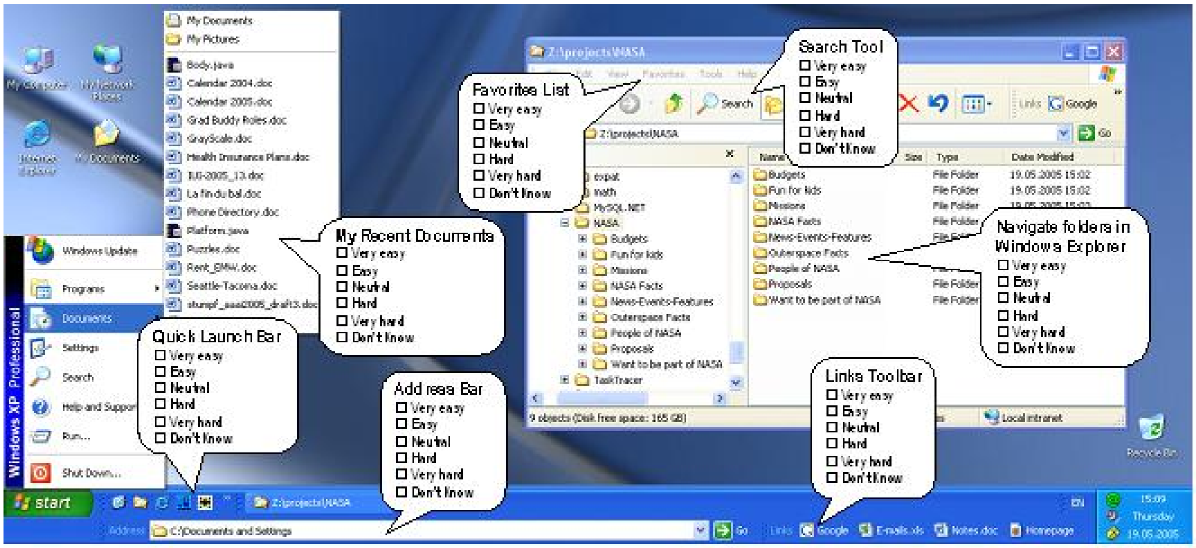
(For example, if you have payroll information about students, would you store it in Payroll folder, students folder, or both, or some other way?)

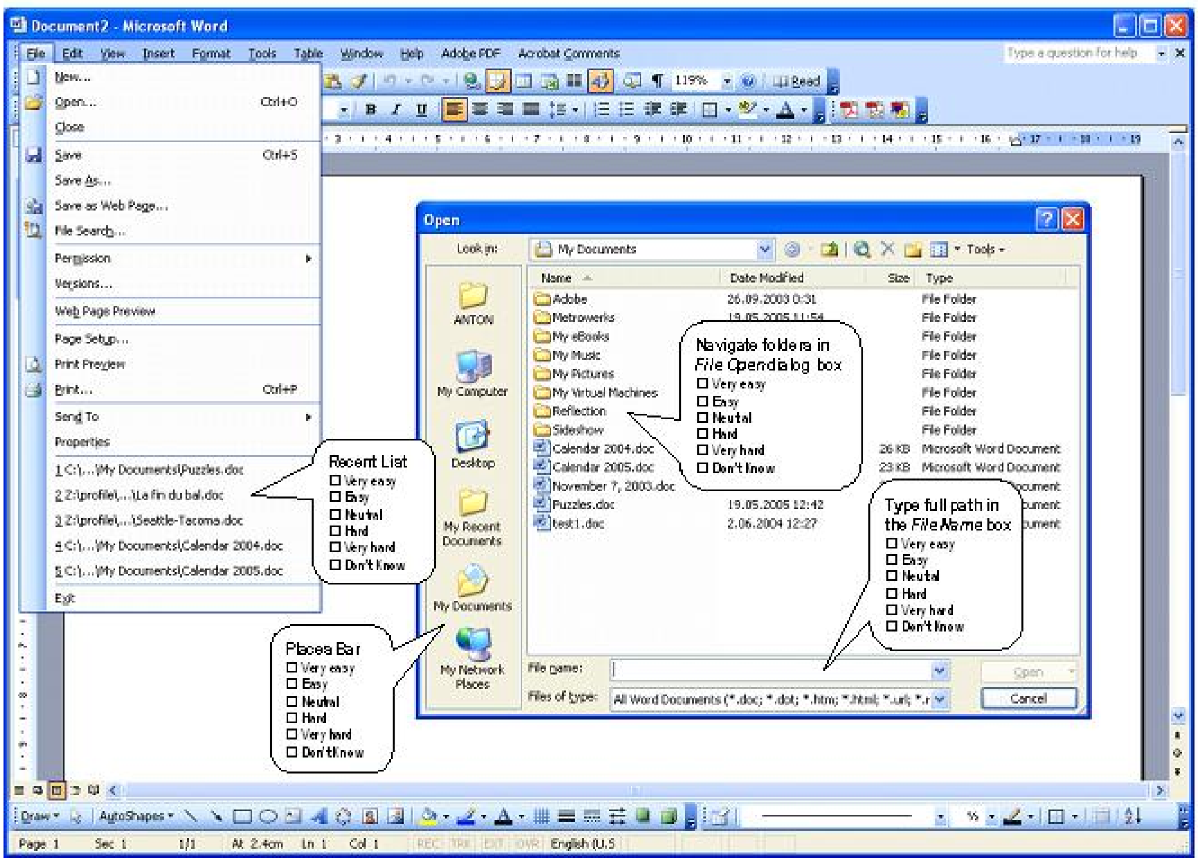
8. Are you familiar with the following ways to access files in MS Windows environment?



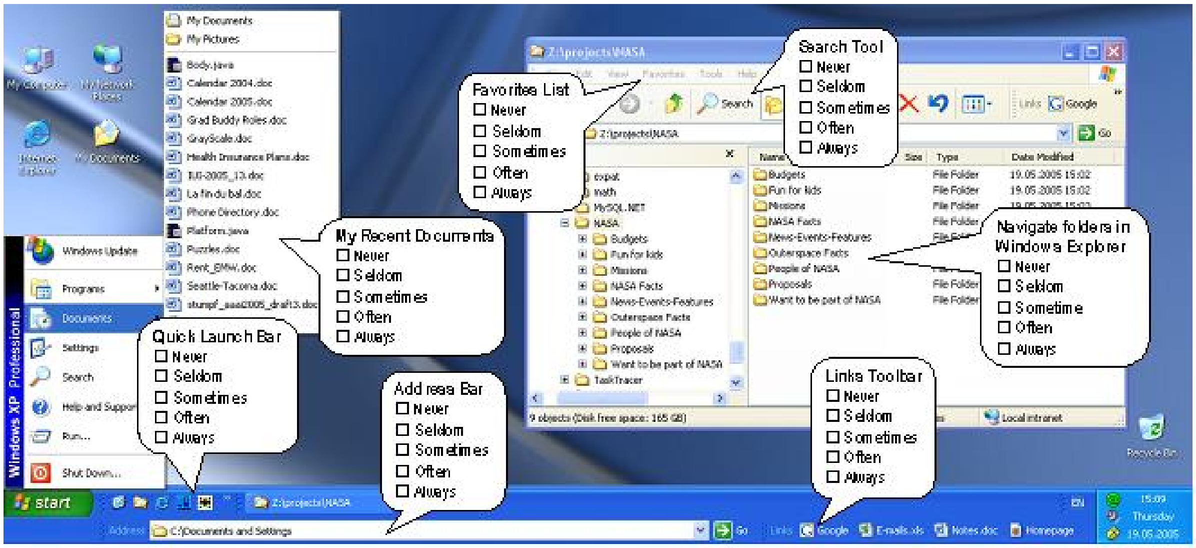


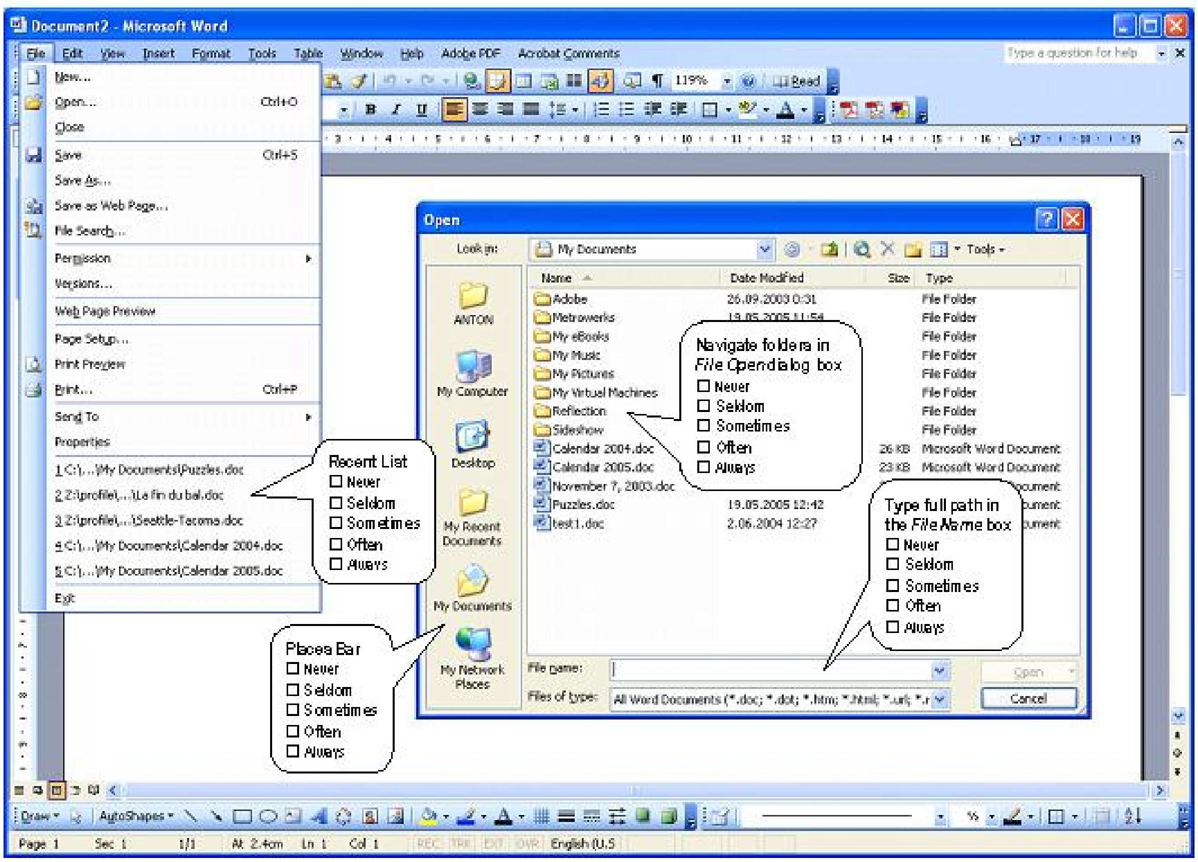
9. How much work do you think it is to find the right files in the following ways (see picture)?





10. How frequently do you use these different ways to get to files in MS Windows environment?





11. Please check any one of the following, considering the interruptions.

	LOW				HIGH
MENTAL DEMAND How much mental and perceptual activity (e.g thinking, deciding, calculating, remembering, looking, searching etc.) was required when you were interrupted?	☐	☐	☐	☐	☐
FRUSTRATION LEVEL How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel when you were interrupted?	☐	☐	☐	☐	☐

THANK YOU